PCT

ORLD INTELLECTUAL PROPERTY ORGANIZAT
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : G10L 9/14

A1

(11) International Publication Number: WO 91/13432

(43) International Publication Date: 5 September 1991 (05.09.91)

(21) International Application Number: PCT/CA90/00381

(22) International Filing Date: 6 November 1990 (06.11.90)

(30) Priority data:
010,830 23 February 1990 (23.02.90) CA

(71) Applicant (*for all designated States except US*): UNIVERSITE DE SHERBROOKE [CA/CA]; Sherbrooke, Quebec J1K 2R1 (CA).

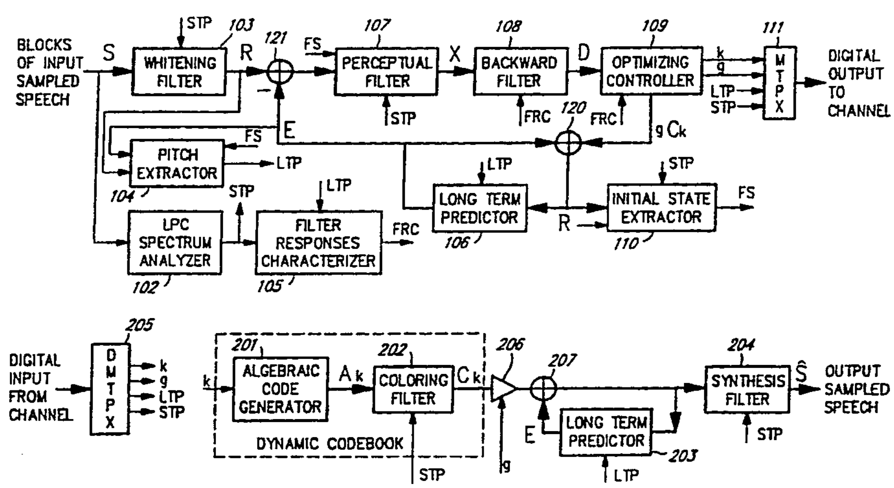
(72) Inventors; and
(75) Inventors/Applicants (*for US only*) : ADOUL, Jean-Pierre [CA/CA]; 2201 Boul. Université, Sherbrooke, Quebec J1K 2P8 (CA). LAFLAMME, Claude [CA/CA]; 390 Rue Farwell, Sherbrooke, Quebec J1J 2S8 (CA).

(74) Agent: GOUDREAU GAGE DUBUC & MARTINEAU WALKER; 3400 The Stock Exchange Tower, P.O. Box 242, Victoria Square, Montreal, Quebec H4Z 1E9 (CA).

(81) Designated States: AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CF (OAPI patent), CG (OAPI patent), CH (European patent), CM (OAPI patent), DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GA (OAPI patent), GB (European patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU (European patent), MC, MG, ML (OAPI patent), MR (OAPI patent), MW, NL (European patent), NO, RO, SD, SE (European patent), SN (OAPI patent), SU, TD (OAPI patent), TG (OAPI patent), US.

Published
With international search report.

(54) Title: DYNAMIC CODEBOOK FOR EFFICIENT SPEECH CODING BASED ON ALGEBRAIC CODES



(57) Abstract

A method of encoding a speech signal is disclosed. This method improves the excitation codebook and search procedure of the conventional Code Excited Linear Prediction (CELP) speech encoders. Use is made of a dynamic codebook (201, 202) based on a combination of two modules: a sparce algebraic code generator (201) associated to a filter (202) having a transfer function varying in time. The generator (201) is a structured codebook with codewords having very few non zero components. The filter (202) shapes the spectral characteristics whereby the resulting excitation codebook (201, 202) exhibits favorable perceptual properties. The search complexity in finding the best codeword is greatly reduced by bringing the search back to the algebraic code domain thereby allowing the sparcity of the algebraic code to speed up the necessary computations.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		

DYNAMIC CODEBOOK FOR EFFICIENT SPEECH CODING BASED ON ALGEBRAIC CODES

BACKGROUND OF THE INVENTION

1. Field of the invention:

The present invention relates to a new technique for digitally encoding and decoding in particular but not exclusively speech signals in view of transmitting and synthesizing these speech signals.

2. Brief description of the prior art:

Efficient digital speech encoding techniques with good subjective quality/bit rate tradeoffs are increasingly in demand for numerous applications such as voice transmission over satellites, land mobile, digital radio or packed network, for voice storage, voice response and secure telephony.

One of the best prior art methods capable of achieving a good quality/bit rate tradeoff is the so called Code Excited Linear Prediction (CELP) technique. In accordance with this method, the speech signal is sampled and converted into successive blocks

of a predetermined number of samples. Each block of samples is synthesized by filtering an appropriate innovation sequence from a codebook, scaled by a gain factor, through two filters having transfer functions varying in time. The first filter is a Long Term Predictor filter (LTP) modeling the pseudoperiodicity of speech, in particular due to pitch, while the second one is a Short Term Predictor filter (STP) modeling the spectral characteristics of the speech signal. The encoding procedure used to determine the parameters necessary to perform this synthesis is an analysis by synthesis technique. At the encoder end, the synthetic output is computed for all candidate innovation sequences from the codebook. The retained codeword is the one corresponding to the synthetic output which is closer to the original speech signal according to a perceptually weighted distortion measure.

The first proposed structured codebooks are called stochastic codebooks. They consist of an actual set of stored sequences of N random samples. More efficient stochastic codebooks propose derivation of a codeword by removing one or more elements from the beginning of the previous codeword and adding one or more new elements at the end thereof. More recently, stochastic codebooks based on linear combinations of a small set of stored basis vectors have greatly reduced the search complexity. Finally, some algebraic structures have also been proposed as excitation codebooks with efficient search procedures.

However, the latter are designed for speed and they lack flexibility in constructing codebooks with good subjective quality characteristics.

OBJECTS OF THE INVENTION

The main object of the present invention is to combine an algebraic codebook and a filter with a transfer function varying in time, to produce a dynamic codebook offering both the speed and memory saving advantages of the above discussed structured codebooks while reducing the computation complexity of the Code Excited Linear Prediction (CELP) technique and enhancing the subjective quality of speech.

SUMMARY OF THE INVENTION

More specifically, in accordance with the present invention, there is provided a method of producing an excitation signal that can be used in synthesizing a sound signal, comprising the steps of generating a codeword signal in response to an index signal associated to this codeword signal, such signal generating step using an algebraic code to generate the codeword signal, and filtering the so generated codeword signal to produce the excitation signal.

Advantageously, the algebraic code is a sparce algebraic code.

The subject invention also relates to a dynamic codebook for producing an excitation signal that can be used in synthesizing a sound signal, comprising means for generating a codeword signal in response to an index signal associated to this codeword signal, which signal generating means using an algebraic code to generate the codeword signal, and means for filtering the so generated codeword signal to produce the excitation signal.

In accordance with a preferred embodiment of the dynamic codebook, the filtering means comprises a coloring filter having a transfer function varying in time to shape the frequency characteristics of the excitation signal so as to damp frequencies perceptually annoying the human ear. This coloring filter comprises an input supplied with linear predictive coding parameters representative of spectral characteristics of the the sound signal to vary the above mentioned transfer function.

In accordance with other aspects of the present invention, there is also provided:

(1) a method of selecting one particular algebraic codeword that can be processed to produce a signal excitation for a synthesis means capable of synthesizing a sound signal, comprising the steps of

(a) whitening the sound signal to be synthesized to generate a residual signal, (b) computing a target signal **X** by processing a difference between the residual signal and a long term prediction component of the signal excitation, (c) backward filtering the target signal to calculate a value **D** of this target signal in the domain of an algebraic code, (d) calculating, for each codeword among a plurality of available algebraic codewords **Ak** expressed in the algebraic code, a target ratio which is function of the value **D**, the codeword **Ak**, and a transfer function **H = D / X** , and (e) selecting the said one particular codeword among the plurality of available algebraic codewords in function of the calculated target ratios.

(2) an encoder for selecting one particular algebraic codeword that can be processed to produce a signal excitation for a synthesis means capable of synthesizing a sound signal, comprising (a) means for whitening the sound signal to be synthesized and thereby generating a residual signal, (b) means for computing a target signal **X** by processing a difference between the residual signal and a long term prediction component of the signal excitation, (c) means for backward filtering the target signal to calculate a value **D** of this target signal in the domain of an algebraic code, (d) means for calculating, for each codeword among a plurality of available algebraic codewords **Ak** expressed in the above mentioned algebraic code, a target ratio which is function of the value **D**, the codeword **Ak**, and a transfer function

$H = D / X$, and (e) means for selecting the said one particular codeword among the plurality of available algebraic codewords in function of the calculated target ratios. In accordance with preferred embodiments of the encoder, the target ratio comprises a numerator given by the expression $P^2(k) = (DA_k^T)^2$ and a denominator given by the expression $\alpha^2 k = \|A_k H^T\|^2$, where A_k and H are under the form of matrix, each codeword A_k is a waveform comprising a small number of non-zero impulses each of which can occupy different positions in the waveform to thereby enable composition of different codewords, the target ratio calculating means comprises means for calculating into a plurality of embedded loops contributions of the non-zero impulses of the considered algebraic codeword to the numerator and denominator and for adding the so calculated contributions to previously calculated sum values of these numerator and denominator, respectively, the embedded loops comprise an inner loop, and the codeword selecting means comprises means for processing in the inner loop the calculated target ratios to determine an optimized target ratio and means for selecting the said one particular algebraic codeword in function of this optimized target ratio.

(3) a method of generating at least one long term prediction parameter related to a sound signal in view of encoding this sound signal, comprising the steps of (a) whitening the sound signal to generate a residual signal, (b) producing a long term prediction component of a signal excitation for a synthesis means

component of a signal excitation for a synthesis means capable of synthesizing the sound signal, which producing step including estimating an unknown portion of the long term prediction component with the residual signal, and (c) calculating the long term prediction parameter in function of the so produced long term prediction component of the signal excitation.

(4) a device for generating at least one long term prediction parameter related to a sound signal in view of encoding this sound signal, comprising (a) means for whitening the sound signal and thereby generating a residual signal, (b) means for producing a long term prediction component of a signal excitation for a synthesis means capable of synthesizing the sound signal, these producing means including means for estimating an unknown portion of the long term prediction component with the residual signal, and (c) means for calculating the long term prediction parameter in function of the so produced long term prediction component of the signal excitation.

The objects, advantages and other features of the present invention will become more apparent upon reading of the following, non restrictive description of a preferred embodiment thereof, given with reference to the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

In the appended drawings:

Figure 1 is a schematic block diagram of the preferred embodiment of an encoding device in accordance with the present invention;

Figure 2 is a schematic block diagram of a decoding device using a dynamic codebook in accordance with the present invention;

Figure 3 is a flow chart showing the sequence of operations performed by the encoding device of Figure 1;

Figure 4 is a flow chart showing the different operations carried out by a pitch extractor of the encoding device of Figure 1, for extracting pitch parameters including a delay T and a pitch gain b; and

Figure 5 is a schematic representation of a plurality of embedded loops used in the computation of optimum codewords and code gains by an optimizing controller of the encoding device of Figure 1.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

Figure 1 is the general block diagram of a speech encoding device in accordance with the present invention. Before being encoded by the device of Figure 1, an analog input speech signal is filtered, typically in the band 200 to 3400 Hz and then sampled at the Nyquist rate (e.g. 8 kHz). The resulting signal comprises a train of samples of varying amplitudes represented by 12 to 16 bits of a digital code. The train of samples is divided into blocks which are each L samples long. In the preferred embodiment of the present invention, L is equal to 60. Each block has therefore a duration of 7.5 ms. The sampled speech signal is encoded on a block by block basis by the encoding device of Figure 1 which is broken down into 10 modules numbered from 102 to 111. The sequence of operation performed by these modules will be described in detail hereinafter with reference to the flow chart of Figure 3 which presents numbered steps. For easy reference, a step number in Figure 3 and the number of the corresponding module in Figure 1 have the same last two digits. Bold letters refer to L-sample-long blocks (i.e. L-component vectors). For instance, **S** stands for the block [S(1), S(2),...S(L)].

<u>Step 301:</u> The next block **S** of L samples is supplied to the encoding device of Figure 1.

<u>Step 302:</u> For each block of L samples of speech signal, a set of Linear Predictive Coding (LPC) parameters, called STP parameters, is produced in accordance with a prior art technique through an LPC spectrum analyser 102. More specifically, the latter analyser 102 models the spectral characteristics of each block S of samples. In the preferred embodiment, the parameters STP comprise a number M=10 of prediction coefficients [a1, a2,...aM]. One can refer to the book by J.D. Markel & A.H. Gray, Jr: "Linear Prediction of Speech" Springer Verlag (1976) to obtain information on representative methods of generating these parameters.

<u>Step 303:</u> The input block S is whitened by a whitening filter 103 having the following transfer function based on the current values of the STP prediction parameters:

$$A(z)=\sum_{i=0}^{M} a_i z^{-i} \qquad (1)$$

where $a_0 = 1$, and z represents the variable of the polynomial A(z).

As illustrated in Figure 1, the filter 103 produces a residual signal R.

Of course, as the processing is performed on a block basis, unless otherwise stated, all the filters are assumed to store their final state for use as initial state in the following block processing.

The purpose of step 304 is to compute the speech periodicity characterized by the Long Term Prediction (LTP) parameters including a delay T and a pitch gain b.

Before further describing step 304, it is useful to explain the structure of the speech decoding device of Figure 2 and understand the principle upon which speech is synthesized.

As shown in Figure 2, a demultiplexer 205 interprets the binary information received from a digital input channel into four types of parameters, namely the parameters STP, LTP, k and g. The current block **s** of speech signal is synthetized on the basis of these four parameters as will be seen hereinafter.

The decoding device of Figure 2 follows the classical structure of the CELP (Code Excited Linear Prediction) technique insofar as modules 201 and 202 are considered as a single entity: the (dynamic) codebook. The codebook is a virtual (i.e. not actually stored) collection of L-sample-long waveforms (codeword) indexed by an integer k. The index k ranges from 0 to NC-1 where NC is the size of the codebook. This size is 4096 in the preferred

embodiment. In the CELP technique, the output speech signal is obtained by first scaling the k^{th} entry of the codebook by the pitch gain g through an amplifier 206. An adder 207 adds the so obtained scaled waveform, gCk, to the output E (the long term prediction component of the signal excitation of a synthesis filter 204) of a long term predictor 203 placed in a feedback loop and having a transfer function B(z) defined as follows:

$$B(z)=bz^{-T} \qquad (2)$$

where b and T are the above defined pitch gain and delay, respectively.

The predictor 203 is a filter having a transfer function influenced by the last received LTP parameters b and T to model the pitch periodicity of speech. It introduces the appropriate pitch gain b and delay of T samples. The composite signal gCk + E constitutes the signal excitation of the sythesis filter 204 which has a transfer function 1/A(z). The filter 204 provides the correct spectrum shaping in accordance with the last received STP parameters. More specifically, the filter 204 models the resonant frequencies (formants) of speech. The output block $\hat{s}$ is the synthesized (sampled) speech signal which can be converted into an analog signal with proper anti-

aliasing filtering in accordance with a technique well known in the art.

In the present invention, the codebook is dynamic; it is not stored but is generated by the two modules 201 and 202. In a first step, an algebraic code generator 201 produces in response to the index k and in accordance with a Sparce Algebraic Code (SAC) a codeword **Ak** formed of a L-sample-long waveform having very few non zero components. In fact, the generator 201 constitutes an inner, structured codebook of size NC. In a second step, the codeword **Ak** from the generator 201 is processed by a coloring filter 202 whose transfer function F(z) varies in time in accordance with the STP parameters. The filter 202 colors, i.e. shapes the frequency characteristics (dynamically controls the frequency) of the output excitation signal Ck so as to damp a priori those frequencies perceptually more annoying to the human ear. The excitation signal Ck, sometimes called the innovation sequence, takes care of whatever part of the original speech signal left unaccounted by either the above defined formant and pitch modelling. In the preferred embodiment of the present invention, the transfer function F(z) is given by the following relationship:

$$F(z)=\frac{A(z\gamma_1^{-1})}{A(z\gamma_2^{-1})} \qquad (3)$$

where γ_1=.7 and γ_2=.85.

There are many ways to design the generator 201. An advantageous method consists of interleaving four single-pulse permutation codes as follows. The codewords Ak are composed of four non zero pulses with fixed amplitudes, namely $S_1=1$, $S_2=-1$, $S_3=1$, and $S_4=-1$. The positions allowed for S_i are of the form $p_i=2i+8m_i-1$, where $m_i=0, 1, 2, \ldots 7$. It should be noted that for $m_3=7$ (or $m_4=7$) the position p_3 (or p_4) falls beyond $L=60$. In such a case, the impulse is simply discarded. The index k is obtained in a straightforward manner using the following relationship:

$$k = 512\, m_1 + 64\, m_2 + 8\, m_3 + m_4 \qquad (4)$$

The resulting Ak-codebook is accordingly composed of 4096 waveforms having only 2 to 4 non zero impulses.

Returning to the encoding procedure, it is useful to discuss briefly the criterion used to select the best excitation signal Ck. This signal must be chosen to minimize, in some ways, the difference $\hat{s} - s$ between the synthesized and original speech signals. In original CELP formulation, the excitation signal Ck is based on a Mean Squared Error (MSE) criteria applied to the error $\Delta = \hat{s}' - s'$, where $\hat{s}'$, respectively s', is $\hat{s}$, respectively s, processed by a

perceptual weighting filter of the form $A(z)/A(z\gamma^{-1})$ where $\gamma = 0.8$ is the perceptual constant. In the present invention, the same criterion is used but the computations are performed in accordance with a backward filtering procedure which is now briefly recalled. One can refer to the article by J.P. Adoul, P. Mabilleau, M. Delprat, & S. Morissette: "Fast CELP coding based on algebraic codes", Proc. IEEE Int'l conference on acoustics speech and signal processing, pp 1957-1960 (April 1987), for more details on this procedure. Backward filtering brings the search back to the Ck-space. The present invention brings the search further back to the Ak-space. This improvement together with the very efficient search method used by controller 109 (Figure 1) and discussed hereinafter enables a tremendous reduction in computation complexity with regard to the conventional approaches.

It should be noted here that the combined transfer function of the filters 103 and 107 (Figure 1) is precisely the same as that of the above mentioned perceptual weighting filter which transforms S into S', that is transforms S into the domain where the MSE criterion can be applied.

<u>Step 304:</u> To carry out this step, a pitch extractor 104 (Figure 1) is used to compute and quantize the LTP parameters , namely the pitch delay T ranging from Tmin to Tmax (20 to 146 samples in the preferred embodiment) and the pitch gain g. Step 304 itself comprises a plurality of steps as illustrated in

Figure 4. Referring now to Figure 4, a target signal **Y** is calculated by filtering (step 402) the residual signal **R** through the perceptual filter 107 with its initial state set (step 401) to the value FS available from an initial state extractor 110. The initial state of the extractor 104 is also set to the value FS as illustrated in Figure 1. The long term prediction component of the signal excitation, E(n), is not known for the current values n = 1, 2, ... The values E(n) for n = 1 to L-Tmin+1 are accordingly estimated using the residual signal **R** available from the filter 103 (step 403). More specifically, E(n) is made equal to R(n) for these values of n. In order to start the search for the best pitch delay T, two variables Max and τ are initialized to 0 and Tmin respectively (step 404). With the initial state set to zero (step 405), the long term prediction part of the signal excitation shifted by the value τ, E(n-τ), is processed by the perceptual filter 107 to obtain the signal **Z**. The crosscorrelation ρ between the signals **Y** and **Z** is then computed using the expression in block 406 of Figure 4. If the crosscorrelation ρ is greater than the variable Max (step 407), the pitch delay T is updated to τ, the variable Max is updated to the value of the crosscorrelation ρ and the pitch energy term α_p equal to $\|Z\|$ is stored (step 410). If τ is smaller than Tmax (step 411), it is incremented by one (step 409) and the search procedure continues. When τ reaches Tmax, the optimum pitch gain b is computed and quantized using the expression b=Max/α_p (step 412).

Step 305: In step 305, a filter responses characterizer 105 (Figure 1) is supplied with the STP and LTP parameters to compute a filter responses characterization FRC for use in the later steps. The FRC information consists of the following three components where n = 1, 2, ... L. It should also be noted that the component f(n) includes the long term prediction loop.

•f(n): impulse response of $F(z)\dfrac{1}{1-bz^{-T}}$ (5a)

•h(n): response of $\dfrac{1}{A(z\gamma^{-1})}$ to f(n) (5b)

with zero initial state.

•u(i,j): autocorrelation of h(n); i.e.:

$$u(i,j)=\sum_{k=1}^{L} h(k-i+1)h(k-j+1) \text{ ; for } 1\leq i\leq L \quad (5c)$$

and $i\leq j\leq L$; $h(n)=0$ for $n<1$

The utility of the FRC information will become obvious upon discussion of the forthcoming steps.

Step 306: The long term predictor 106 is supplied with the signal excitation E + gCk to compute the component E of this excitation contributed by the long term prediction (parameters LTP) using the proper pitch delay T and gain b. The predictor 106 has the same transfer function as the long term predictor 203 of Figure 2.

Step 307: In this step, the initial state of the perceptual filter 107 is set to the value FS supplied by the initial state extractor 110. The difference R-E calculated by a subtractor 121 (Figure 1) is then supplied to the perceptual filter 107 to obtain at the output of the latter filter a target block signal X. As illustrated in Figure 1, the STP parameters are applied to the filter 107 to vary its transfer function in relation to these parameters. Basically, X = S' - P where P represents the contribution of the long term prediction (LTP) including "ringing" from the past excitations. The MSE criterion which applies to Δ can now be stated in the following matrix notations.

$$\min_k \|\Delta\|^2 = \min_k \|S' - \hat{S}\|^2 = \min_k \|S' - [P - gA_k H^T]\|^2 \qquad (6)$$

$$= \min_k \|X - gA_k H^T\|^2$$

where H accounts for the global filter transfer function $F(z)/(1-B(z))A(z\gamma^{-1})$. It is an L x L lower triangular Toeplitz matrix formed from the h(n) response.

<u>Step 308:</u> This is the backward filtering step performed by the filter 108 of Figure 1. Setting to zero the derivative of the above equation (6) with respect to the code gain g yields to the optimum gain as follows:

$$\frac{\partial\|\Delta\|^2}{\partial g}=0$$

$$g=\frac{X(A_kH^T)^T}{\|A_kH^T\|^2} \tag{7}$$

With this value for g the minimization becomes:

$$\begin{aligned}\min_k\|\Delta\|^2 &= \min_k\left\{\|X\|^2-\frac{\left(X(A_kH^T)^T\right)^2}{\|A_kH^T\|^2}\right\}\\ &= \max_k\frac{\left(X(A_kH^T)^T\right)^2}{\|A_kH^T\|^2} \qquad (8)\\ &= \max_k\frac{\left((XH)A_k{}^T\right)^2}{\alpha^2{}_k}=\max_k\frac{\left(DA_k{}^T\right)^2}{\alpha^2{}_k}\end{aligned}$$

where $D=(XH)$ and $\alpha^2{}_k=\|A_kH^T\|^2$.

In step 308, the backward filtered target signal $D=(XH)$ is computed. The term "backward filtering" for this operation comes from the interpretation of (XH) as the filtering of time-reversed X.

Step 309: In this step performed by the optimizing controller 109 of Figure 1, equation (8) is optimized by computing the ratio $(DAk^T/\alpha k)^2 = P^2k/\alpha^2k$ for each sparce algebraic codeword Ak. The denominator is given by the expression:

$$\alpha^2_k = \left\| A_k H^T \right\|^2 = A_k H^T H A_k{}^T = A_k U A_k{}^T \qquad (9)$$

where U is the Toeplitz matrix of the autocorrelations defined in equation (5c). Calling S(i) and p(i) respectively the amplitude and position of the ith non zero impulse (i = 1, 2, ...N), the numerator and (squared) denominator simplify to the following:

$$DA_k{}^T = \sum_{i=1}^{N} S(i) D(p_i) \qquad (10a)$$

$$\alpha^2_k = \sum_{i=1}^{N} S^2(i) U(p_i, p_i) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} S(i) S(j) U(p_i, p_j) \qquad (10b)$$

where $P(N) = DAk^T$

A very fast procedure for calculating the above defined ratio for each codeword Ak is described in Figure 5 as a set of N embedded computation loops, N being the number of non zero impulses in the codewords. The quantities $S^2(i)$ and $SS(i,j) = S(i)S(j)$, for i=1, 2, ... N and $i < j \leq N$ are pre-stored for maximum speed. Prior to the computations, the values for P^2_{opt} and α^2_{opt} are initialized to zero and some large number, respectively. As can be seen in Figure 5, partial sums of the numerator and denominator are calculated in each one of the outer and inner loops, while in the inner loop the largest ratio $P^2(N)/\alpha^2(N)$ is retained as the ratio P^2_{opt}/α^2_{opt}. The calculating procedure is believed to be otherwise self-explanatory from Figure 5. When the N embedded loops are completed, the code gain is computed as $g = P_{opt} / {}^{2}_{opt}$ (cf. equation (7)). The gain is then quantized, the index k is computed from stored impulse positions using the expression (4), and the L components of the scaled optimum code gCk are computed as follows:

$$gC_k(n) = g\sum_{i=1}^{N} f(n-p_i) \quad ;1 \leq n \leq L \qquad (11)$$
$$\text{with } f(n) = 0 \text{ ; for } n < 1$$

Step 310: The global signal excitation signal E + gCk is computed by an adder 120 (Figure 1). The initial state extractor module 110, constituted by a

perceptual filter with a transfer function $1/A(z\gamma^{-1})$ varying in relation to the STP parameters, subtracts from the residual signal **R** the signal excitation signal **E** + gCk for the sole purpose of obtaining the final filter state FS for use as initial state in filter 107 and module 104.

<u>Step 311:</u> The set of four parameters STP, LTP, k and g are converted into the proper digital channel format by a multiplexer 111 completing the procedure for encoding a block *S* of samples of speech signal.

Accordingly, the present invention provides a fully quantized Algebraic Code Excited Linear Prediction (ACELP) vocoder giving near toll quality at rates ranging from 4 to 16 kbits. This is achieved through the use of the above described dynamic codebook and associated fast search algorithm.

The drastic complexity reduction that the present invention offers when compared to the prior art techniques comes from the fact that the search procedure can be brought back to **A**k-code space by a modification of the so called backward filtering formulation. In this approach the search reduces to finding the index k for which the ratio $|DAk^T|/\alpha k$ is the largest. In this ratio, **A**k is a fixed target signal and αk is an energy term the computation of which can be done with very few operations by codeword when N, the number of non zero components of the codeword **A**k, is small.

Although a preferred embodiment of the present invention has been described in detail hereinabove, this embodiment can be modified at will, within the scope of the appended claims, without departing from the nature and spirit of the invention. As an example, many types of algebraic codes can be chosen to achieve the same goal of reducing the search complexity while many types of coloring filters can be used. Also the invention is not limited to the treatment of a speech signal; other types of sound signal can be processed. Such modifications, which retain the basic principle of combining an algebraic code generator with a coloring filter, are obviously within the scope of the subject invention.

The embodiments of the invention in which an exclusive property or privilege is claimed are defined as follows.

1. A method of producing an excitation signal that can be used in synthesizing a sound signal, comprising the steps of:

generating a codeword signal in response to an index signal associated to said codeword signal, said signal generating step using an algebraic code to generate the said codeword signal; and

filtering the so generated codeword signal to produce said excitation signal.

2. A method as defined in claim 1, in which the algebraic code is a sparce algebraic code.

3. A method as defined in claim 1, wherein the excitation signal has frequency characteristics, and wherein said filtering step comprises processing the codeword signal through a coloring filter having a transfer function varying in time to thereby shape the frequency characteristics of the excitation signal so as to damp frequencies perceptually annoying the human ear.

4. A method as defined in claim 3, in which the transfer function of the coloring filter is varied in relation to linear predictive coding parameters representative of spectral characteristics of the said sound signal.

5. A dynamic codebook for producing an excitation signal that can be used in synthesizing a sound signal, comprising:

means for generating a codeword signal in response to an index signal associated to said codeword signal, said signal generating means using an algebraic code to generate the said codeword signal; and

means for filtering the so generated codeword signal to produce said excitation signal.

6. A codebook as defined in claim 5, in which the algebraic code is a sparce algebraic code.

7. A codebook as defined in claim 5, wherein the excitation signal has frequency characteristics, and wherein said filtering means comprises a coloring filter having a transfer function varying in time to shape the frequency characteristics of the excitation signal so as to damp frequencies perceptually annoying the human ear.

8. A codebook as defined in claim 7, in which the coloring filter comprises an input supplied with linear predictive coding parameters representative of spectral characteristics of the said sound signal to vary the said transfer function.

9. A method of selecting one particular algebraic codeword that can be processed to produce a signal excitation for a synthesis means capable of synthesizing a sound signal, comprising the steps of:

whitening said sound signal to be synthesized to generate a residual signal;

computing a target signal **X** by processing a difference between the said residual signal and a long term prediction component of said signal excitation;

backward filtering the target signal to calculate a value **D** of the said target signal in the domain of an algebraic code;

calculating, for each codeword among a plurality of available algebraic codewords **Ak** expressed in the said algebraic code, a target ratio which is function of the value **D**, the codeword **Ak**, and a transfer function $H = D / X$; and

selecting the said one particular codeword among said plurality of available algebraic codewords in function of the calculated target ratios.

10. The selecting method of claim 9, in which said target ratio comprises a numerator given by

the expression $P^2(k) = (DAk^T)^2$ and a denominator given by the expression $\alpha^2 k = \| AkH^T \|^2$, where Ak and H are under the form of matrix.

11. The selecting method of claim 10, wherein each codeword Ak is a waveform comprising a small number of non-zero impulses each of which can occupy different positions in the waveform to thereby enable composition of different codewords.

12. The selecting method of claim 11, in which said target ratio calculating step uses a calculating procedure including embedded loops in which are calculated contributions of the non-zero impulses of the considered algebraic codeword to the said numerator and denominator and in which the so calculated contributions are added to previously calculated sum values of said numerator and denominator, respectively.

13. The selecting method of claim 12, wherein the embedded loops comprise an inner loop, and wherein the said codeword selecting step comprises the steps of:

processing in the inner loop the said calculated target ratios to determine an optimized target ratio; and

selecting the said one particular algebraic codeword in function of said optimized target ratio.

14. The selecting method of claim 9, wherein the said codeword selecting step comprises the steps of:

processing the said calculated target ratios to determine an optimized target ratio; and

selecting the said one particular algebraic codeword in function of said optimized target ratio.

15. An encoder for selecting one particular algebraic codeword that can be processed to produce a signal excitation for a synthesis means capable of synthesizing a sound signal, comprising:

means for whitening said sound signal to be synthesized and thereby generating a residual signal;

means for computing a target signal X by processing a difference between the said residual signal and a long term prediction component of said signal excitation;

means for backward filtering the target signal to calculate a value D of the said target signal in the domain of an algebraic code;

means for calculating, for each codeword among a plurality of available algebraic codewords Ak expressed in the said algebraic code, a

target ratio which is function of the value D, the codeword Ak, and a transfer function $H = D / X$; and means for selecting the said one particular codeword among said plurality of available algebraic codewords in function of the calculated target ratios.

16. The encoder of claim 15, in which said target ratio comprises a numerator given by the expression $P^2(k) = (DAk^T)^2$ and a denominator given by the expression $\alpha^2 k = \| AkH^T \|^2$, where Ak and H are under the form of matrix.

17. The encoder of claim 16, wherein each codeword Ak is a waveform comprising a small number of non-zero impulses each of which can occupy different positions in the waveform to thereby enable composition of different codewords.

18. The encoder of claim 17, in which said target ratio calculating means comprises means for calculating into a plurality of embedded loops contributions of the non-zero impulses of the considered algebraic codeword to the said numerator and denominator and for adding the so calculated contributions to previously calculated sum values of said numerator and denominator, respectively.

19. The encoder of claim 18, wherein the embedded loops comprise an inner loop, and wherein the said codeword selecting means comprises:

means for processing in the inner loop the said calculated target ratios to determine an optimized target ratio; and

means for selecting the said one particular algebraic codeword in function of said optimized target ratio.

20. The encoder of claim 15, wherein the said codeword selecting means comprises:

means for processing the said calculated target ratios to determine an optimized target ratio; and

means for selecting the said one particular algebraic codeword in function of said optimized target ratio.

21. A method of generating at least one long term prediction parameter related to a sound signal in view of encoding the said sound signal, comprising the steps of:

whitening said sound signal to generate a residual signal;

producing a long term prediction component of a signal excitation for a synthesis means capable of synthesizing the said sound signal, said producing step including estimating an unknown portion

of the long term prediction component with the said residual signal; and

calculating the said at least on long term prediction parameter in function of the so produced long term prediction component of said signal excitation.

22. A device for generating at least one long term prediction parameter related to a sound signal in view of encoding the said sound signal, comprising:

means for whitening said sound signal and thereby generating a residual signal;

means for producing a long term prediction component of a signal excitation for a synthesis means capable of synthesizing the said sound signal, said producing means including means for estimating an unknown portion of the long term prediction component with the said residual signal; and

means for calculating the said at least one long term prediction parameter in function of the so produced long term prediction component of said signal excitation.

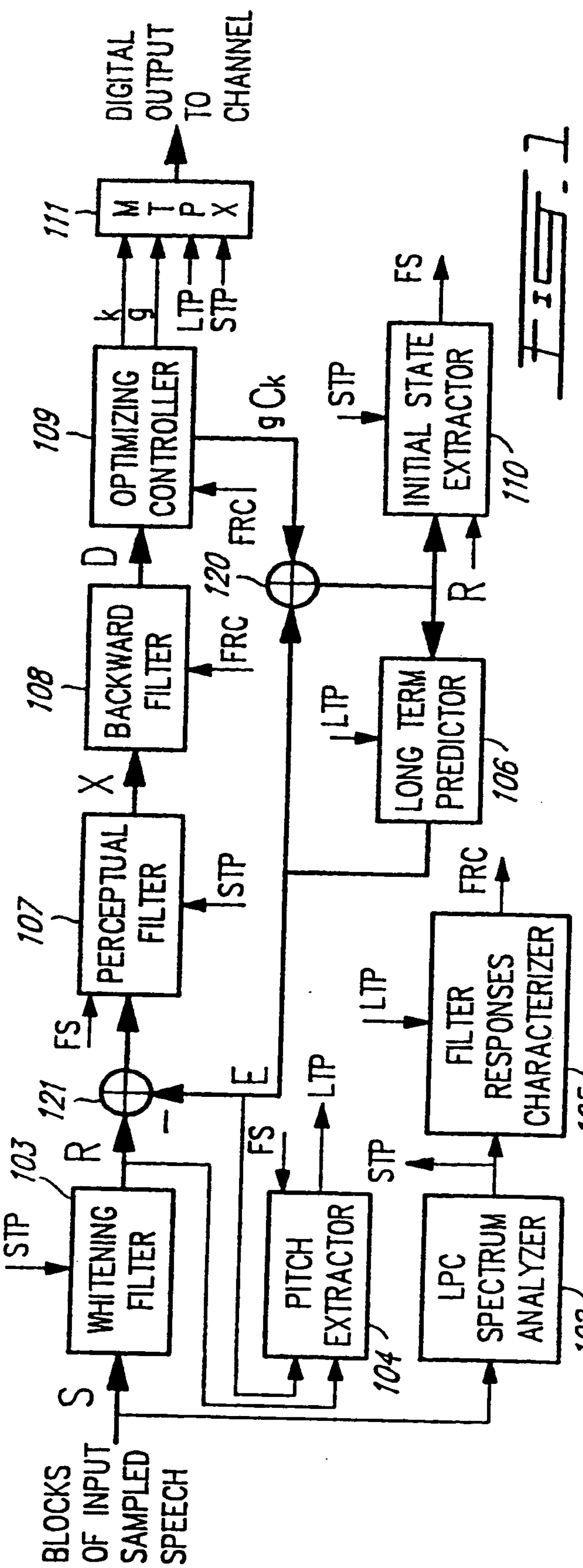

BLOCKS OF INPUT SAMPLED SPEECH
S
STP
WHITENING FILTER
103
R
121
FS
PERCEPTUAL FILTER
107
STP
X
BACKWARD FILTER
108
FRC
D
OPTIMIZING CONTROLLER
109
FRC
k
g
LTP
STP
MTPX
111
DIGITAL OUTPUT TO CHANNEL
gCk
120
R
E
FS
PITCH EXTRACTOR
104
LTP
LONG TERM PREDICTOR
106
INITIAL STATE EXTRACTOR
110
FS
LPC SPECTRUM ANALYZER
102
STP
FILTER RESPONSES CHARACTERIZER
105
FRC

FIG. 1

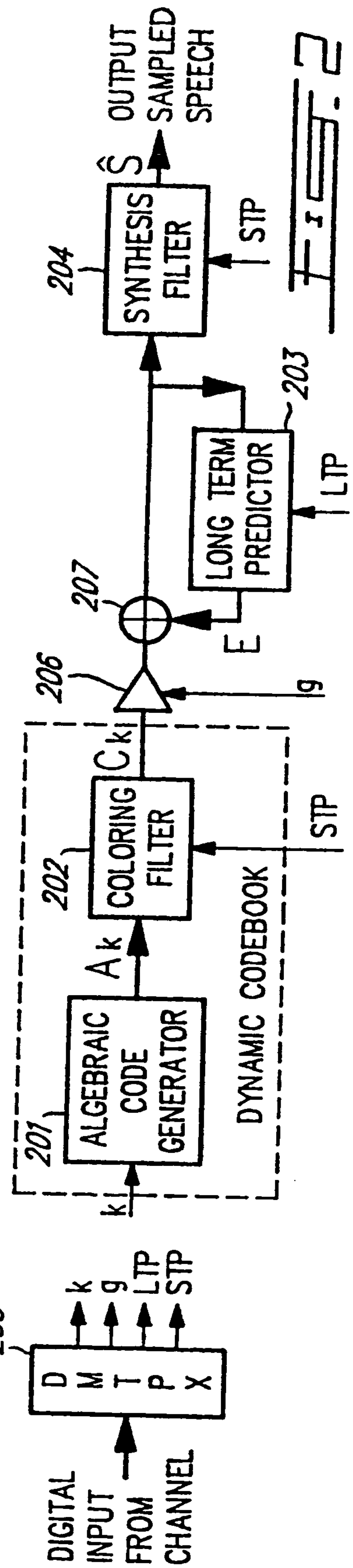

DIGITAL INPUT FROM CHANNEL
DMTPX
205
k
g
LTP
STP
ALGEBRAIC CODE GENERATOR
201
Ak
COLORING FILTER
202
STP
Ck
DYNAMIC CODEBOOK
206
g
207
E
LONG TERM PREDICTOR
203
LTP
SYNTHESIS FILTER
204
STP
Ŝ
OUTPUT SAMPLED SPEECH

FIG. 2

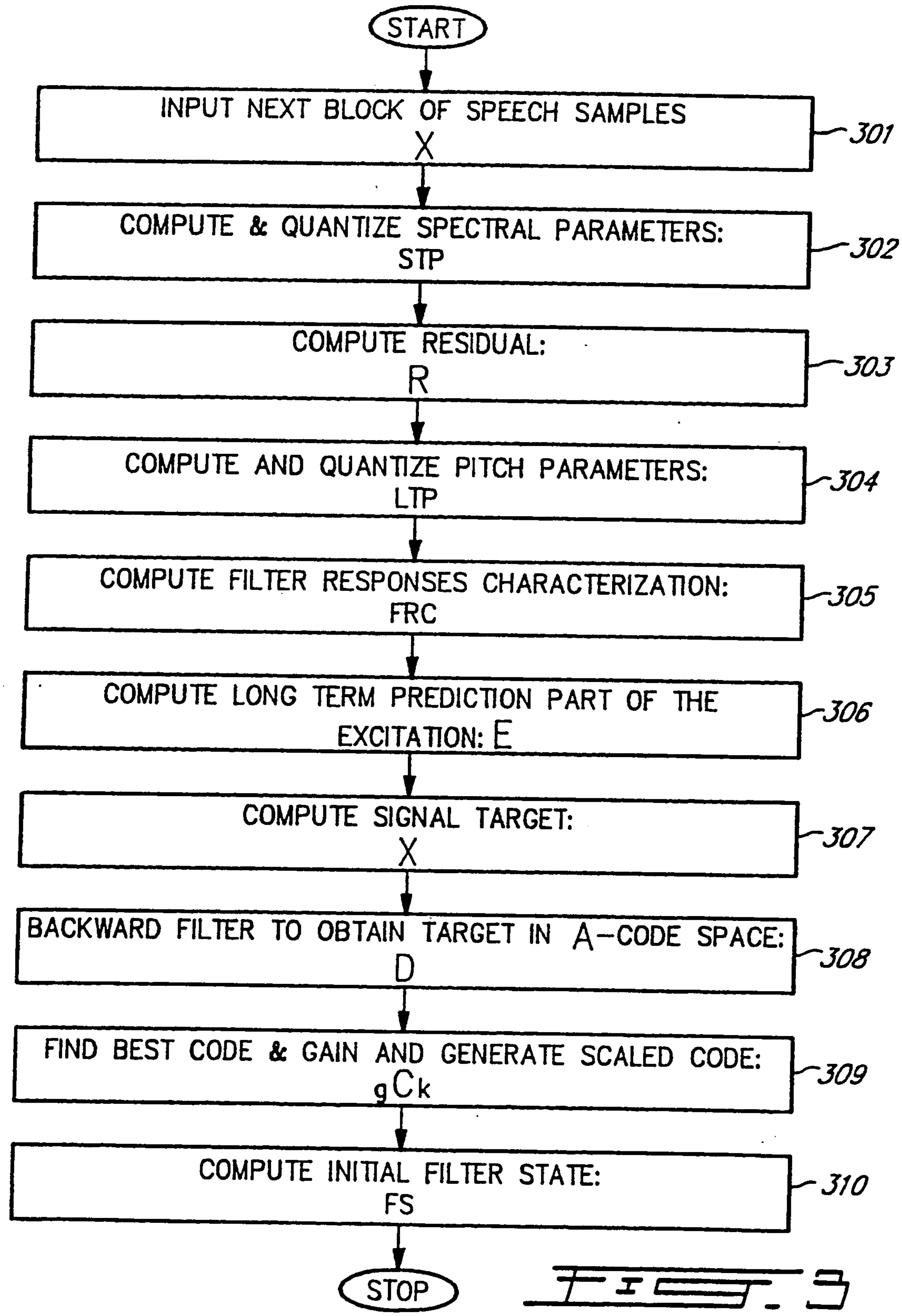

START
INPUT NEXT BLOCK OF SPEECH SAMPLES
X
301
COMPUTE & QUANTIZE SPECTRAL PARAMETERS:
STP
302
COMPUTE RESIDUAL:
R
303
COMPUTE AND QUANTIZE PITCH PARAMETERS:
LTP
304
COMPUTE FILTER RESPONSES CHARACTERIZATION:
FRC
305
COMPUTE LONG TERM PREDICTION PART OF THE EXCITATION: E
306
COMPUTE SIGNAL TARGET:
X
307
BACKWARD FILTER TO OBTAIN TARGET IN A-CODE SPACE:
D
308
FIND BEST CODE & GAIN AND GENERATE SCALED CODE:
gCk
309
COMPUTE INITIAL FILTER STATE:
FS
310
STOP

FIG. 3

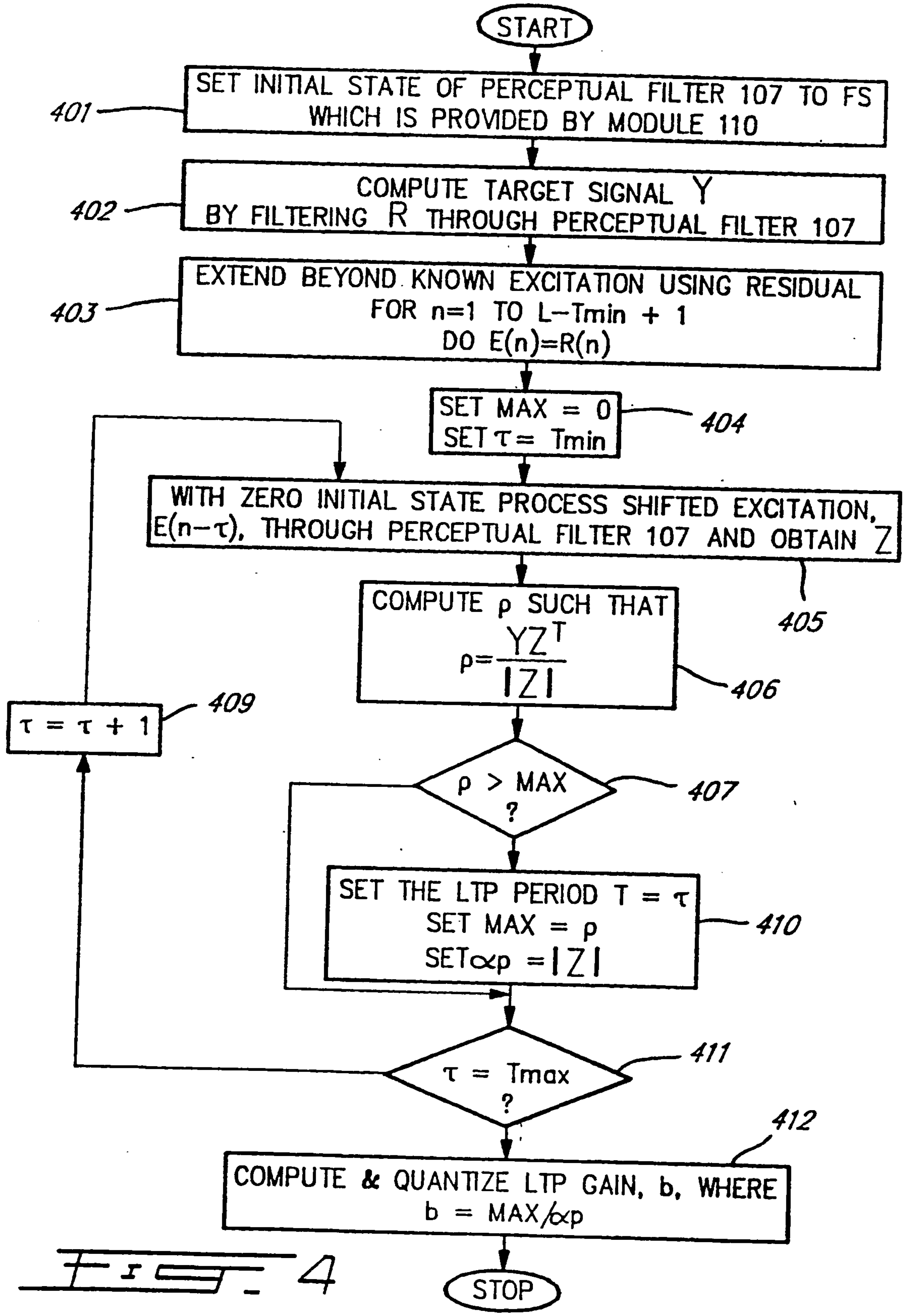

START
SET INITIAL STATE OF PERCEPTUAL FILTER 107 TO FS WHICH IS PROVIDED BY MODULE 110
401
COMPUTE TARGET SIGNAL Y BY FILTERING R THROUGH PERCEPTUAL FILTER 107
402
EXTEND BEYOND KNOWN EXCITATION USING RESIDUAL FOR n=1 TO L−Tmin + 1 DO E(n)=R(n)
403
SET MAX = 0 SET τ = Tmin
404
WITH ZERO INITIAL STATE PROCESS SHIFTED EXCITATION, E(n−τ), THROUGH PERCEPTUAL FILTER 107 AND OBTAIN Z
405
COMPUTE ρ SUCH THAT ρ = YZ^T / |Z|
406
τ = τ + 1
409
ρ > MAX ?
407
SET THE LTP PERIOD T = τ SET MAX = ρ SET αp = |Z|
410
τ = Tmax ?
411
COMPUTE & QUANTIZE LTP GAIN, b, WHERE b = MAX/αp
412
STOP

FIG. 4

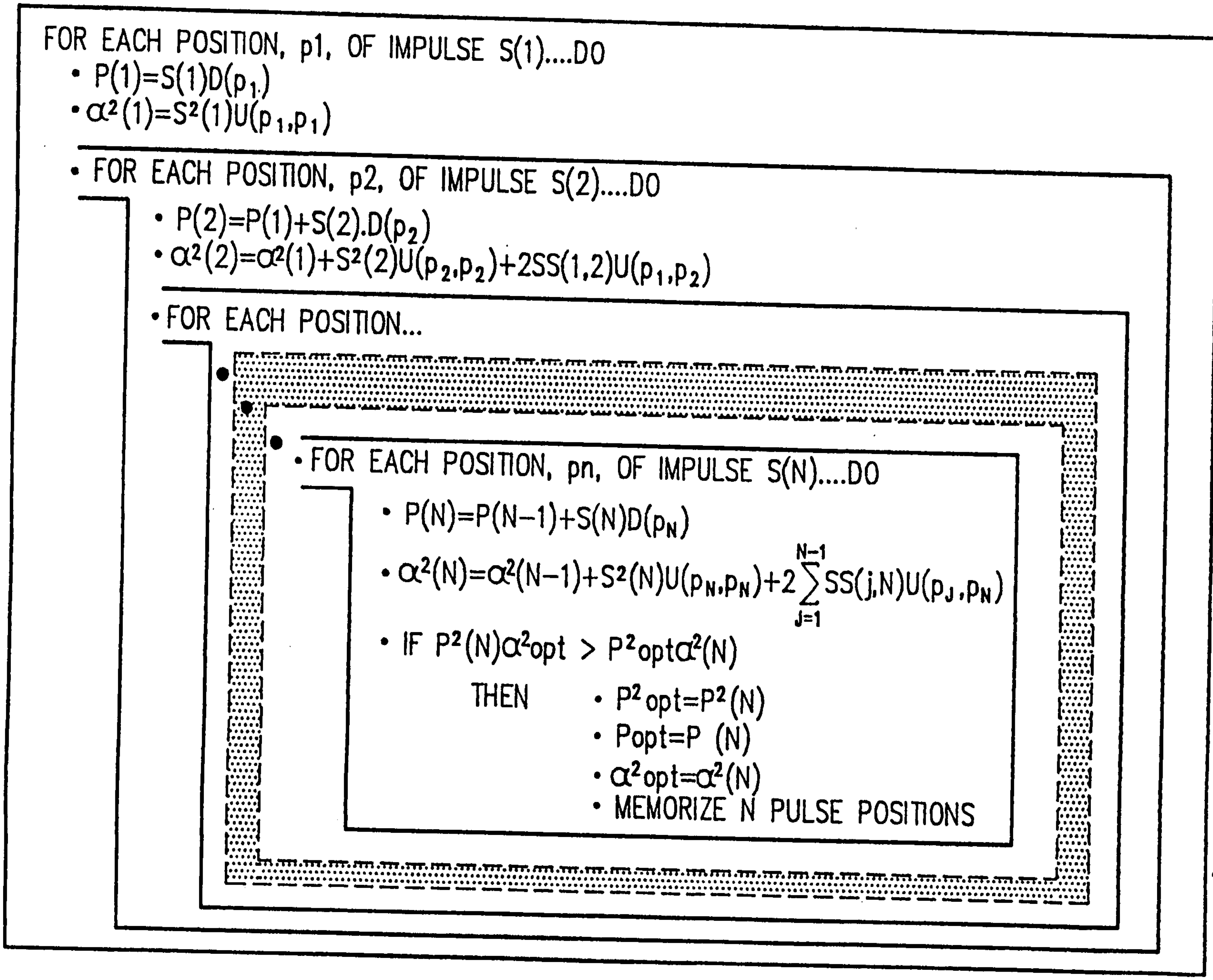

FOR EACH POSITION, p1, OF IMPULSE S(1)....DO
• P(1)=S(1)D(p1)
• α²(1)=S²(1)U(p1,p1)
• FOR EACH POSITION, p2, OF IMPULSE S(2)....DO
• P(2)=P(1)+S(2).D(p2)
• α²(2)=α²(1)+S²(2)U(p2,p2)+2SS(1,2)U(p1,p2)
• FOR EACH POSITION...
• FOR EACH POSITION, pn, OF IMPULSE S(N)....DO
• P(N)=P(N−1)+S(N)D(pN)
• α²(N)=α²(N−1)+S²(N)U(pN,pN)+2 Σ(J=1 to N−1) SS(j,N)U(pJ,pN)
• IF P²(N)α²opt > P²optα²(N)
THEN
• P²opt=P²(N)
• Popt=P (N)
• α²opt=α²(N)
• MEMORIZE N PULSE POSITIONS

FIG. 5

INTERNATIONAL SEARCH REPORT

International Application No PCT/CA 90/00381

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC^5: G 10 L 9/14

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System	Classification Symbols
IPC^5	G 10 L 9/14

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category [*]	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
A	ICASSP 86, IEEE-IECEJ-ASJ International Conference on Acoustics, Speech, and Signal Processing, 7-11 April 1986, Tokyo, Japan, IEEE (New York, US), A. le Guyader et al.: "A robust 16 kbits/s vector adaptive predictive coder for mobile communications", pages 857-860 see figure 1 --	1,3-5,7,8
A	IEEE Global Telecommunications Conference & Exhibition, 28 November - 1 December 1988, Hollywood, Florida, US, volume 1, IEEE, (New York, US), F.F. Tzeng: "Multipulse excitation codebook design and fast search methods for CELP speech coding", pages 590-594 --	1,5

./.

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
14th February 1991	13. 03. 91
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	N. Toribio Nuria TORIBIO

Form PCT/ISA/210 (second sheet) (January 1985)

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, [11] with indication, where appropriate, of the relevant passages	Relevant to Claim No.
A	ICASSP 87, 1987 International Conference on Acoustics, Speech, and Signal Processing, 6-9 April 1987, Dallas, Texas, US, volume 4, IEEE, (New York, US), J.-P. Adoul et al.: "A comparison of some algebraic structures for CELP coding of speech", pages 1953-1956	2,6
